```
=> d his

     (FILE 'HOME' ENTERED AT 18:42:47 ON 20 MAR 2004)
     FILE 'CA' ENTERED AT 18:43:01 ON 20 MAR 2004
L1 251301 S MASS SPEC? OR MS MS
L2    367 S L1 AND(MINE# OR MINING)
L3   1292 S L1 AND((DATA OR SPECTRA# OR SPECTRUM)(3A)(REDUCT? OR DATABASE OR
          SEARCHING OR ALGORITHM OR INDEX? OR SCORE# OR SCORING))
L4    337 S L1 AND((DATABASE OR DATA BASE)(3A)(REDUCT? OR SEARCHING OR
          ALGORTHM OR INDEX? OR SCORE# OR SCORING))
L5    589 S L2-4 AND(COMPUTER OR MICROPROCESSOR OR ALGORITHM OR AUTOMAT?)
L6      7 S L1 AND(MINE# OR MINING)(10A)(COMPUTER OR MICROPROCESSOR OR
          ALGORITHM OR SOFTWARE OR AUTOMAT?)
L7    324 S L1 AND((DATA OR SPECTRA# OR SPECTRUM)(3A)(REDUCT? OR DATABASE OR
          SEARCHING OR ALGORITHM OR INDEX? OR SCORE# OR SCORING))(10A)(
          COMPUTER OR MICROPROCESSOR OR ALGORITHM OR SOFTWARE OR AUTOMAT?)
L8     83 S L1 AND((DATABASE OR DATA BASE)(3A)(REDUCT? OR SEARCHING OR
          ALGORITHM OR INDEX? OR SCORE# OR SCORING))(10A)(COMPUTER OR
          MICROPROCESSOR OR ALGORITHM OR SOFTWARE OR AUTOMAT?)
L9    376 S L6-8
L10   219 S L5 NOT L9
L11   273 S L9 NOT PY>2000
L12    14 S L9 NOT L11 AND PATENT/DT
L13   147 S L10 NOT PY>2000
L14   324 S L11-13 AND(COMPUTER? OR DATABASE OR ALGORITHM OR INTERPRET? OR
          RETRIEV?)/TI,IT

=> d bib,ab 1-324 l14

L14  ANSWER 10 OF 324  CA  COPYRIGHT 2004 ACS on STN
AN   136:45371  CA
TI   Method and system for mining mass spectral data
IN   Hansen, Beau; Liebler, Daniel C.; Mason, Daniel E.; Jones, Juliet A.
PA   University of Arizona, Board of Regents, USA
SO   PCT Int. Appl., 63 pp.
PI   WO 2001097251     A1   20011220       WO 2001-US18798  20010612
PRAI US 2000-210981P   P    20000612
AB   Methods for mining mass spectra are described which entail specifying
     spectral characteristics (e.g., product ion, loss ion, and/or ion
     series) of the mass spectra to mine; specifying a relationship between
     the spectral characteristics; searching the mass spectra for portions
     of the mass spectra which match the spectral characteristics based on
     the relationship; and assigning scores to the portions of the mass
     spectra to indicate a degree of correlation between the portions of the
     mass spectra and the spectral characteristics.  The mass spectra may be
     obtained by dissocn.(e.g., collision-induced dissocn.) or full-scan.
     Systems for mining mass spectra and computer programs including a
     graphical user interface for carrying out the mining are also
     described.

L14  ANSWER 16 OF 324  CA  COPYRIGHT 2004 ACS on STN
```

AN 134:258759 CA
TI New deconvolution method for electrospray ionization **mass spectrometry**
AU Kato, Hiroshi; Ishihara, Morio; Nakata, Munetaka
CS JEOL Ltd., Akishima, Tokyo, 196-8558, Japan
SO Journal of the Mass Spectrometry Society of Japan (2000), 48(6), 373-379
AB A new method is proposed for elimination of artifacts appearing in deconvolution of electrospray ionization **mass spectra**, where two **algorithms**, a partial correlation method (PCM) and a sub-harmonic artifact removal filter (SHARF), are used. In addn. to the elimination of artifacts, the former algorithm removes influence of singly charged ions generated from contamination in a sample, while the latter algorithm removes influence of background noises and baseline offsets in a measured spectrum. The proposed method results in supplying the deconvoluted spectra free from artifacts with good signal-to-noise ratios and without distortion on peak shapes. Applications to some bio-mols. lead to the conclusion that our method is esp. useful for analyses of their mixt. samples, which show complicated **mass spectra**.

L14 ANSWER 17 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 134:190282 CA
TI Searching sequence **databases** via de novo peptide sequencing by tandem **mass spectrometry**
AU Johnson, Richard S.; Taylor, J. Alex
CS USA
SO Methods in Molecular Biology (Totowa, New Jersey) (2000), 146(Mass Spectrometry of Proteins and Peptides), 41-61
AB Three computer programs have been written that together provide an alternative approach to searching sequence databases using tandem **mass spectra**. The first (Lutefisk97) performs a de novo sequence interpretation and provides, as output, a short list of candidate sequences. It is important to note that manual or computer interpretations of low-energy collision-induced decompn. (CD) data from tandem **mass spectra** of peptides are bound to yield multiple sequence candidates, and it is often impossible to distinguish the correct sequence from the incorrect ones. Frequently, the variations between sequence possibilities are minor and involve inversions of dipeptides, swapping dipeptides of the same mass, or swapping single amino acids with dipeptides of the same mass. To deal with these multiple yet similar sequence candidates, a second computer program (CIDentify) was written. CIDentify is a version of the FASTA homol.-based search program that was modified to accommodate the ambiguous sequencing results obtained by tandem **mass spectrometry**. The third program, CIDentify Result Compiler, compiles the CIDentify output for peptides derived from the same protein, and produces a list of database sequences that are ranked according to the no. and quality of individual matches found by CIDentify. It is hereby described how to obtain, implement, and run these programs.

L14 ANSWER 18 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 134:155013 CA

TI Development of a Windows NT-based dynamic SIMS **software** program for instrument control and **data reduction** across different instrument platforms
AU McNitt, P. J.; Hagen, J. J.; Martel, D. J.; Register, R. A.
CS Physical Electronics, Redwood City, CA, 94063, USA
SO Secondary Ion Mass Spectrometry, SIMS XII, Proceedings of the International Conference on Secondary Ion Mass Spectrometry, 12th, Brussels, Belgium, Sept. 5-10, 1999 (2000), Meeting Date 1999, 355-358. Editor(s): Benninghoven, Alfred. Publisher: Elsevier Science B.V., Amsterdam, Neth.
AB New software, running under Windows NT, is described which provides instrument control and data redn. for magnetic sector, time of flight, and quadrupole SIMS instruments. Data redn. tools and techniques were developed based on the previous generation of software.

L14 ANSWER 20 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 134:27296 CA
TI An expert system for protein identification using **mass spectrometric** information combined with **database searching**
IN Zhang, Wenzhu; Chait, Brian T.; Fenyo, David; Tang, Chao
PA Rockefeller University, USA; Proteometrics, LLC
SO PCT Int. Appl., 64 pp.
PI WO 2000073787 A1 20001207 WO 2000-US14809 20000526
PRAI US 1999-136267P P 19990527
AB A method is disclosed for detg. the probability that an exptl. biol. mol. is a biol. mol. described in a **database** given exptl. mass **data** and background information. A 30 kDa SDS-PAGE protein band from a Saccharomyces cerevisiae nuclear ext. was in-gel digested with trypsin and the peptides were analyzed by delayed-extn. reflectron MALDI-TOF. The thirty-five monoisotopic masses that were derived were submitted to ProFound as well as other search parameters (e.g., taxonomy, protein mass range, etc.) in order to identify the protein. The top 4 protein candidates were listed as ranked by normalized probability.

L14 ANSWER 24 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 133:292955 CA
TI Identifying the proteome: software tools
AU Fenyo, David
CS ProteoMetrics, LLC, New York, NY, 10018, USA
SO Current Opinion in Biotechnology (2000), 11(4), 391-395
AB A review with 44 refs. The interest in proteomics has recently increased dramatically and proteomic methods are now applied to many problems in cell biol. The method of choice in proteomics for identifying and characterizing proteins is **mass spectrometry** combined with **database searching**. Software tools have been improved to increase the sensitivity of protein identification and methods for evaluating the search results have been incorporated.

L14 ANSWER 27 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 133:101575 CA
TI Method for screening peptide fragment ion **mass spectra** prior to

database searching

AU Moore, R. E.; Young, M. K.; Lee, T. D.

CS Beckman Research Institute of the City of Hope, Duarte, CA, USA

SO Journal of the American Society for Mass Spectrometry (2000), 11(5), 422-426

AB A methodol. is described for screening fragment ion spectra of peptides prior to database searching for protein identification. A software routine written in the Perl programming language was used to analyze data from previous Sequest database searches and develop a set of statistical descriptors that could be used to identify spectra not likely to yield useful results in a database search. A second Perl program used an evolutionary algorithm to optimize the criteria for each statistical descriptor and generate a formula for detg. spectral quality. This formula was used by a third Perl program to screen data sets from four independent liq. chromatog. tandem **mass spectrometry** runs. On the av., use of the screening program reduced the time required for a database search by 1/2 with little loss of useful information from the database search results.

L14 ANSWER 28 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 133:54508 CA

TI Methods and materials for peptide-based DNA sequence determination and analysis

IN Jarvik, Jonathan W.

PA Sequel Genetics, Inc., USA

SO PCT Int. Appl., 59 pp.

PI	WO 2000036414	A1	20000622	WO 1999-US30104	19991216
	US 2002155445	A1	20021024	US 2001-788268	20010216
PRAI	US 1998-112351P	P	19981216		

AB A nucleic acid fragment of interest is incorporated into a hybrid artificial gene and expressed in one or more reading frames to produce one or more hybrid polypeptides. The polypeptides are examd. with respect to one or more phys. parameters, such as mass or amino acid compn. The obsd. parameter values are used to search a data set of predicted parameter values generated by hypothetical translation of a larger ref. nucleic acid sequence so as to det. whether or not the fragment is contained within the ref. sequence, and, if it is contained therein, to det. its sequence and/or coding capacity. The method can be applied to the identification of genetic mutations and polymorphism, phenotypes, genotyping, disease diagnosis or prognosis. **Computer** based database, storage medium, and programs for **searching** and analyzing the **data** sets, are also claimed. Sequence of the nucleic acid fragments corresponding to the human nucleolin gene and alpha complementing factor of beta-galactosidase gene were correctly identified by the method utilizing MALDI-TOF **mass spectrometry** anal. of the peptides. Specific genetic mutations and polymorphism were also identified. A computer program for calcg. the mass shifts arising from single nucleotide substitutions was developed.

L14 ANSWER 29 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 132:305437 CA

TI An **Algorithm** for Automated Bacterial Identification Using Matrix-Assisted Laser Desorption/Ionization **Mass Spectrometry**

AU Jarman, Kristin H.; Cebula, Sharon T.; Saenz, Adam J.; Petersen, Catherine E.; Valentine, Nancy B.; Kingsley, Mark T.; Wahl, Karen L.

CS Pacific Northwest National Laboratory, Richland, WA, 99352, USA

SO Analytical Chemistry (2000), 72(6), 1217-1223

AB An algorithm for bacterial identification using matrix-assisted laser desorption/ionization (MALDI) **mass spectrometry** is being developed. This **mass spectral** fingerprint comparison **algorithm** is fully **automated** and statistically based, providing objective anal. of samples to be identified. Based on extn. of ref. fingerprint ions from test spectra, this approach should lend itself well to real-world applications where samples are likely to be impure. This algorithm is illustrated using a blind study. In the study, MALDI-MS fingerprints for Bacillus atrophaeus ATCC 49337, Bacillus cereus ATCC 14579T, Escherichia coli ATCC 33694, Pantoea agglomerans ATCC 33243, and Pseudomonas putida F1 are collected and form a ref. library. The identification of test samples contg. one or more ref. bacteria, potentially mixed with one species not in the library (Shewanella alga BrY), is performed by comparison to the ref. library with a calcd. degree of assocn. Out of 60 samples, no false positives are present, and the correct identification rate is 75%. Missed identifications are largely due to a weak B. cereus signal in the bacterial mixts. Potential modifications to the algorithm are presented and result in a higher than 90% correct identification rate for the blind study data, suggesting that this approach has the potential for reliable and accurate automated data anal. of MALDI-MS.

L14 ANSWER 31 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 132:276130 CA

TI A Statistical Basis for Testing the Significance of **Mass Spectrometric** Protein Identification Results

AU Eriksson, Jan; Chait, Brian T.; Fenyoe, David

CS The Rockefeller University, New York, NY, 10021, USA

SO Analytical Chemistry (2000), 72(5), 999-1005

AB A method for testing the significance of **mass spectrometric** (MS) protein identification results is presented. MS proteolytic peptide mapping and genome database searching provide a rapid, sensitive, and potentially accurate means for identifying proteins. **Database** search **algorithms** detect the matching between proteolytic peptide masses from an MS peptide map and theor. proteolytic peptide masses of the proteins in a genome database. The no. of masses that matches is used to compute a score, S, for each protein, and the protein that yields the best score is assumed as the identification result. There is a risk of obtaining a false result, because masses detd. by MS are not unique; i.e., each mass in a peptide map can match randomly one or several proteins in a genome database. A false result is obtained when the score, S, due to random matching cannot be discerned from the score due to matching with a real protein in the sample. We therefore introduce the frequency function, f(S), for false (random) identification results as a basis for testing at what significance level, α, one can reject a

null hypothesis, H0: "the result is false". The significance is tested by comparing an exptl. score, SE, with a crit. score, SC, required for a significant result at the level α. If SE ≥ SC, H0 is rejected. f(S) and SC were obtained by simulations utilizing random tryptic peptide maps generated from a genome database. The crit. score, SC, was studied as a function of the no. of masses in the peptide map, the mass accuracy, the degree of incomplete enzymic cleavage, the protein mass range, and the size of the genome. With SC known for a variety of exptl. constraints, significance testing can be fully **automated** and integrated with **database searching software** used for protein identification.

L14 ANSWER 32 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 132:148728 CA
TI Improving protein identification from peptide mass fingerprinting through a parameterized multi-level scoring **algorithm** and an optimized peak detection
AU Gras, Robin; Muller, Markus; Gasteiger, Elisabeth; Gay, Steven; Binz, Pierre-Alain; Bienvenut, William; Hoogland, Christine; Sanchez, Jean-Charles; Bairoch, Amos; Hochstrasser, Denis F.; Appel, Ron D.
CS University Medical Center, Swiss Institute of Bioinformatics, Geneva, CH-1211/4, Switz.
SO Electrophoresis (1999), 20(18), 3535-3550
AB We have developed a new **algorithm** to identify proteins by means of peptide mass fingerprinting. Starting from the matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) spectra and environmental data such as species, isoelec. point and mol. wt., as well as chem. modifications or no. of missed cleavages of a protein, the program performs a fully **automated** identification of the protein. The first step is a peak detection **algorithm**, which allows precise and fast detn. of peptide masses, even if the peaks are of low intensity or they overlap. In the second step the masses and environmental data are used by the identification **algorithm** to search in protein sequence databases (SWISS-PROT and/or TrEMBL) for protein entries that match the input data. Consequently, a list of candidate proteins is selected from the **database**, and a **score** calcn. provides a ranking according to the quality of the match. To define the most discriminating scoring calcn. we analyzed the resp. role of each parameter in two directions. The first one is based on filtering and exploratory effects, while the second direction focuses on the levels where the parameters intervene in the identification process. Thus, according to our anal., all input parameters contribute to the score, however with different wts. Since it is difficult to est. the wts. in advance, they have been computed with a generic **algorithm**, using a training set of 91 protein spectra with their environmental data. We tested the resulting scoring calcn. on a test set of ten proteins and compared the identification results with those of other peptide mass fingerprinting programs.

L14 ANSWER 33 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 132:148648 CA
TI Probability-based protein identification by searching sequence

databases using **mass spectrometry** data

AU Perkins, David N.; Pappin, Darryl J. C.; Creasy, David M.; Cottrell, John S.

CS Imperial Cancer Research Fund, London, WC2A 3PX, UK

SO Electrophoresis (1999), 20(18), 3551-3567

AB Several **algorithms** have been described in the literature for protein identification by **searching** a sequence **database** using **mass spectrometry** data. In some approaches, the exptl. data are peptide mol. wts. from the digestion of a protein by an enzyme. Other approaches use tandem **mass spectrometry** (**MS/MS**) data from one or more peptides. Still others combine mass data with amino acid sequence data. We present results from a new computer program, Mascot, which integrates all three types of search. The scoring algorithm is probability based, which has a no. of advantages: (i) A simple rule can be used to judge whether a result is significant or not. This is particularly useful in guarding against false positives. (ii) Scores can be compared with those from other types of search, such as sequence homol. (iii) Search parameters can be readily optimized by iteration. The strengths and limitations of probability-based scoring are discussed, particularly in the context of high throughput, fully automated protein identification.

L14 ANSWER 34 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 132:148647 CA

TI Reliable automatic protein identification from matrix-assisted laser desorption/ionization **mass spectrometric** peptide fingerprints

AU Berndt, Peter; Hobohm, Uwe; Langen, Hanno

CS Gene Technologies, Basel, Switz.

SO Electrophoresis (1999), 20(18), 3521-3526

AB Matrix-assisted laser desorption/ionization (MALDI) **mass spectrometry** of protein samples from two-dimensional (2-D) gels in conjunction with protein sequence database searches is frequently used to identify proteins. Moreover, the automatic anal. of complete 2-D gels with hundreds and even thousands of protein spots ("proteome anal.") is possible, without human intervention, with the availability of highly accurate **mass spectrometry** instruments, and high-throughput facilities for prepn. and handling of protein samples from 2-D gels. However, the lack of software for precise automatic anal. and annotation of **mass spectra**, as well as software for in-batch sequence database queries, is increasingly becoming a significant bottleneck for the proteomics work flow. In the present paper we outline an **algorithm** for reliable, accurate, and **automatic** evaluation of **mass spectrometric data** and **database** searches. We show here that simply selecting from the sequence database the protein that has the most matching fragment masses often leads to false-pos. results. Reliable protein identification is dependent on several parameters: the accuracy of fragment mass detn., the no. of masses submitted for query, the mass distribution of query masses, the no. of masses matching between sample and database protein, the size of the sequence database, and the kind and no. of modifications considered. Using these parameters, we derive a simple statistical estn. that can be used to calc. the probability of true-pos. protein identification.

L14 ANSWER 35 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 132:75579 CA
TI De Novo peptide sequencing via tandem **mass spectrometry**
AU Dancik, Vlado; Addona, Theresa A.; Clauser, Karl R.; Vath, James E.; Pevzner, Pavel A.
CS Millennium Pharmaceuticals, Cambridge, MA, USA
SO Journal of Computational Biology (1999), 6(3/4), 327-342
AB Peptide sequencing via tandem **mass spectrometry** (**MS/MS**) is one of the most powerful tools in proteomics for identifying proteins. Because complete genome sequences are accumulating rapidly, the recent trend in interpretation of **MS/MS** spectra has been database search. However, de novo **MS/MS** spectral interpretation remains an open problem typically involving manual interpretation by expert **mass spectrometrists**. We have developed a new algorithm, SHERENGA, for de novo interpretation that automatically learns fragment ion types and intensity thresholds from a collection of test spectra generated from any type of **mass spectrometer**. The test data are used to construct optimal path scoring in the graph representations of **MS/MS** spectra. A ranked list of high scoring paths corresponds to potential peptide sequences. SHERENGA is most useful for interpreting sequences of peptides resulting from unknown proteins and for validating the results of **database algorithms** in fully **automated**, high-throughput peptide sequencing.

L14 ANSWER 37 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 132:10500 CA
TI A new **spectral interpretation algorithm** for protein sequencing using tandem **mass spectroscopy**
IN Dancik, Vladimir
PA Millennium Pharmaceuticals, Inc., USA
SO PCT Int. Appl., 28 pp.
PI WO 9962930 A2 19991209 WO 1999-US12221 19990602
PRAI US 1998-87785P P 19980603
AB A new algorithm, SHERENGA, for de novo spectral interpretation is described that automatically learns fragment ion-types and intensity thresholds from a collection of test spectra generated from any type of **mass spectrometer**. The algorithm employs a graph theory approach. The test data is used to construct optimal path scoring in the graph representations of tandem **mass spectra**. A ranked list of high scoring paths corresponds to potential peptide sequences. SHERENGA is most useful for interpreting sequences of peptides resulting from unknown proteins not yet encountered in genome sequencing, and leveraging text based pattern matching for homol. matching to known proteins. The algorithm also serves as a powerful adjunct for validating the results of **database** matching **algorithms** in fully **automated**, high-throughput peptide sequencing.

L14 ANSWER 38 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 132:7397 CA
TI On-line acquisition, analysis, and e-mailing of high-resolution exact-mass electron impact/chemical ionization **mass spectrometry** data

acquired using an automated direct probe
AU Huang, N.; Siegel, M. M.; Muenster, H.; Weissenberg, K.
CS Lederle Laboratories, Wyeth-Ayerst Research, Pearl River, NY, USA
SO Journal of the American Society for Mass Spectrometry (1999), 10(11), 1212-1216
AB A complete automation package was developed for data acquisition, processing, interpreting, and e-mailing of high-resoln. exact-mass electron impact (EI) and chem. ionization (CI) **mass spectrometry** data. A com. high performance magnetic sector **mass spectrometer** equipped with a com. programmable robotic direct probe was used. The software package contains modules that **automatically** performs all the functions necessary for **data redn.** and reporting. In sequential order, these functions include downloading of sample information from a corporate database, creation of a sample list, acquisition of high-resoln. exact-mass data, processing of the data, generation of an exact-mass report, e-mailing of the results to the requesting chemists, and finally shutting down of the instrument. The performance of the system was evaluated with nearly 500 samples. The system is reliable and robust with a small av. systematic mass error of -0.47 mmu and a std. deviation of 1.61 mmu.

L14 ANSWER 39 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 132:7396 CA
TI Automated data massaging, **interpretation**, and e-mailing modules for high throughput open access **mass spectrometry**
AU Tong, H.; Bell, D.; Tabei, K.; Siegel, M. M.
CS Lederle Laboratories, Wyeth-Ayerst Research, Pearl River, NY, USA
SO Journal of the American Society for Mass Spectrometry (1999), 10(11), 1174-1187
AB Hardware components and software modules were configured to enhance the automation, efficiency, and reliability of a com. open access atm. pressure ionization **mass spectrometry** (API/MS) system for flow injection anal. The data massaging module is a versatile package for data manipulation/redn. which is initialized upon detecting the end of data acquisition and can function in parallel during the data acquisition of the next sample. The data interpretation module compares the ions in the acquired **mass spectrum** with the predicted mol. adduct ions in different charge states, as well as the predicted isotopic distributions, possible artifact, polymer/cluster, byproduct/fragmentation ions, and then uses the results to score the quality of the spectrum. The e-mailing module transmits the spectrum and interpretation report to the desktop computer of the submitting chemist where the spectrum can be displayed and the report viewed. A scheme is also presented for the automated interpretation of an API **mass spectrum** for the detn. of the most likely mol. wts. of the components present in an unknown sample. Related flow diagrams, algorithms, and applications are illustrated.

L14 ANSWER 42 OF 324 CA COPYRIGHT 2004 ACS on
AN 131:254577 CA
TI Windowed mass selection method: a new **data** processing **algorithm** for

liquid chromatography-**mass spectrometry** data

AU Fleming, Cliona M.; Kowalski, Bruce R.; Apffel, Alex; Hancock, William S.

CS Department of Chemistry, Laboratory for Chemometrics, University of Washington, Seattle, WA, 98195, USA

SO Journal of Chromatography, A (1999), 849(1), 71-85

AB A no. of preprocessing methods are tested on liq. chromatog.-**mass spectrometry** (LC-MS) peptide map data, to det. the best and most efficient way to improve the signal to noise ratio in the data, esp. at low analyte concns. Three methods are investigated, including an algorithm named "sequential paired covariance" (SPC), which was recently reported. An improvement to this algorithm is also reported here. This new, improved method, named the "windowed mass selection method" (WMSM), is shown to effectively eliminate random noise that occurs in the data. This method is shown to be particularly useful in improving signal to noise ratios in both chromatog. and **mass spectra** for data acquired in peptide mapping of recombinant DNA derived proteins.

L14 ANSWER 49 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 130:241670 CA

TI Time compressed gas chromatography/**mass spectrometry**

AU Gankin, Yuriy; Robbat, Albert, Jr.

CS Ion Signature Technology, Inc., Cambridge, MA, USA

SO AT-PROCESS (1999), Volume Date 1998-1999, 4(1,2), 56-62

AB Novel MS interpretation algorithms, which allow for an increase in the overall speed of gas chromatog./**mass spectrometric** anal. of complex environmental samples, are presented. The data anal. system developed using these algorithms virtually eliminates the need for extensive sample prepn. and minimizes chromatog. sepn. times. When coupled with a new thermal desorption sample introduction system, these algorithms provide electron capture detection level sensitivity in the presence of highly concd. sample interferents. Quant. results are presented for an oil contaminated mixt. of polychlorinated biphenyls, chlorinated pesticides and polycyclic arom. hydrocarbons analyzed in less than seven minutes without sample cleanup. The data quality (signal-to-noise ratio, accuracy, and precision) was equal to that produced by std. lab. instruments and methods where each compd. family is typically analyzed sep. and with much longer run times.

L14 ANSWER 51 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 130:145711 CA

TI Spectrum recovery from discrete detector arrays-correction for nonuniformity

AU Birkinshaw, K.

CS Department of Physics, University of Wales Aberystwyth, Ceredigion, SY23 3BZ, UK

SO International Journal of Mass Spectrometry (1998), 181, 159-165

AB In arrays of detectors there is always a degree of nonuniformity. The causes of nonuniformity are outlined for the case of an array of independent detectors and an algorithm is presented that enables

nonuniformity correction and recovery of the spectrum incident on the detector from the measured spectrum under the conditions given.

L14 ANSWER 52 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 130:49374 CA
TI Characterization of Serine and Threonine Phosphorylation Sites in β-Elimination/Ethanethiol Addition-Modified Proteins by Electrospray Tandem **Mass Spectrometry** and **Database Searching**
AU Jaffe, Howard; Veeranna; Pant, Harish C.
CS LNC-NINDS Protein/Peptide Sequencing Facility National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, 20892, USA
SO Biochemistry (1998), 37(46), 16211-16224
AB A new method for the characterization of serine and threonine phosphorylation sites in proteins has been developed. After modification of a phosphoprotein by β-elimination/ethanethiol addn. and conversion of phosphoserine and phosphothreonine residues to S-ethylcysteinyl or β-methyl-S-ethylcysteinyl residues, the modified protein was subjected to proteolytic digestion. Resulting digests were analyzed by a combination of microbore liq. chromatog., electrospray ionization tandem (**MS/MS**) ion trap **mass spectrometry** and **database searching** to identify original phosphorylated residues. The **computer** program utilized (SEQUEST) is capable of identifying peptides and modified residues from uninterpreted **MS/MS** spectra, and using this method, all of the five known phosphorylation sites in bovine β-casein were identified. Application of the method to multiply phosphorylated human high mol. wt. neurofilament protein (NF-H) resulted in the identification of 21 peptides and their modified residues and hence, the in vivo phosphorylation sites. These included 26 KSP and 1 KTP site, all of which occur in the KSP repeat C-terminal tail domain (residues 502-823). One site at residue 518 was previously uncharacterized. A novel non-KSP serine at residue 421 near the KLLEGEE region in a IPFSLPE motif was characterized as phosphorylated (or glycosylated). The 27 characterized phosphorylation sites occur at S/TP residues in the following motifs: KSPVKEE, KSPAEAK, KSPEKEE, KSPAEVK, KSPEKAK, KSPPEAK, KSPVKAE, and KTPAKEE. On the basis of kinase consensus sequences, all of these motifs, including the previously unreported KTPAKEE motif, can be phosphorylated by proline-directed kinases. Advantages of the new method vis-a-vis our previously reported method [Jaffe, H., Veeranna, Shetty, K. T., and Pant, H. C. (1998) Biochem. 37, 3931-3940] include (i) prodn. of diastereomers eluting at different retention times increased the chances of peptide identification, (ii) increased hydrophobicity and hence retention time of the modified peptides, (iii) facilitation of pos. ion prodn., and (i.v.) increased susceptibility to tryptic digestion as a result of conversion of neg. charged phosphorylated residues to neutral S-ethylcysteine or β-methyl-S-ethylcysteine residues.

L14 ANSWER 55 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 129:14031 CA
TI **Database** searching using **mass spectrometry** data
AU Yates, John R., III
CS Dep. Molecular Biotechnol., Univ. Washington, Seattle, WA, 98185, USA
SO Electrophoresis (1998), 19(6), 893-900
AB A review with 48 refs. Large-scale DNA sequencing is creating a sequence infrastructure of great benefit to protein biochem. Concurrent with the application of large-scale DNA sequencing to whole genome anal., **mass spectrometry** has attained the capability to rapidly, and with remarkable sensitivity, det. wts. and amino acid sequences of peptides. Computer algorithms were developed to use the 2 different types of data generated by **mass spectrometers** to search sequence databases. When a protein is digested with a site-specific protease, the mol. wts. of the resulting collection of peptides, the mass map or fingerprint, can be detd. using **mass spectrometry**. The mol. wts. of the set of peptides derived from the digestion of a protein can then be used to identify the protein. Several different approaches were developed. Protein identification using peptide mass mapping is an effective technique when studying organisms with completed genomes. A 2nd method is based on the use of data created by tandem **mass spectrometers**. Tandem **mass spectra** contain highly specific information in the fragmentation pattern as well as sequence information. This information was used to search databases of translated protein sequences as well as nucleotide databases such as expressed sequence tag (EST) sequences. The ability to search nucleotide databases is an advantage when analyzing data obtained from organisms whose genomes are not yet completed, but a large amt. of expressed gene sequence is available (e.g., human and mouse). Furthermore, a strength of using tandem **mass spectra** to search databases is the ability to identify proteins present in fairly complex mixts.

L14 ANSWER 58 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 128:250175 CA
TI New **Computer** Aided Methods for Revealing Structural Features of Unknown Compounds Using Low Resolution **Mass Spectra**
AU Lebedev, Konstantin S.; Cabrol-Bass, Daniel
CS Institute of Organic Chemistry, Siberian Branch of Russian Academy of Science, Novosibirsk, 630090, Russia
SO Journal of Chemical Information and Computer Sciences (1998), 38(3), 410-419
AB Two new computer methods designed to reveal structural features of unknown compds. by low resoln. **mass spectra** are presented. Both methods use the results of a spectral similarity search in a **mass spectral** database. The 1st one proceeds by intersecting selected structures to find maximal common substructures, while the 2nd proceeds by decompg. these structures to derive fragments following a model of primary fragmentation of org. mols. Reliability of the revealed fragments is estd. by comparing an unknown compd.'s spectrum with the computed spectral images of each fragment. The usefulness and limitations of the two proposed methods are estd. by using a set of test examples. In many cases the two methods are complementary,

whereas overall, the 2nd looks more promising both for revealing large structural fragments and for generation of candidate structures, because the fragments revealed have only one or two free valences and rarely overlap one another.

L14 ANSWER 59 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 128:240895 CA
TI Characterization of the Phosphorylation Sites of Human High Molecular Weight Neurofilament Protein by Electrospray Ionization Tandem **Mass Spectrometry** and **Database Searching**
AU Jaffe, Howard; Veeranna; Shetty, K. T.; Pant, Harish C.
CS Protein/Peptide Sequencing Facility and Laboratory of Neurochemistry National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, 20892, USA
SO Biochemistry (1998), 37(11), 3931-3940
AB Hyperphosphorylated high mol. wt. neurofilament protein (NF-H) exhibits extensive phosphorylation on lysine-serine-proline (KSP) repeats in the C-terminal domain of the mol. Specific phosphorylation sites in human NF-H were identified by proteolytic digestion and anal. of the resulting digests by a combination of microbore liq. chromatog., electrospray ionization tandem (**MS/MS**) ion trap **mass spectrometry**, and **database searching**. The **computer** programs utilized (PEPSEARCH and SEQUEST) are capable of identifying peptides and phosphorylation sites from uninterpreted **MS/MS** spectra, and by use of these methods, 27 phosphopeptides and their phosphorylated residues were identified. On the basis of these phosphopeptides, 38 phosphorylation sites in human NF-H were characterized. These include 33 KSP, lysine-threonine-proline (KTP) or arginine-serine-proline (RSP) sites and four unphosphorylated sites, all of which occur in the KSP repeat domain (residues 502-823); and one threonine phosphorylation site obsd. in a KVPTPEK motif. Six KSP sites were not characterized because of the failure to isolate and identify corresponding phosphopeptides. Heterogeneity in serine and threonine phosphorylation was obsd. at three sites or deduced to occur at three sites on the basis of enzyme specificity. As a result of the phosphorylated motifs identified (KSPAKEE, KSPVKEE, KS/TPEKAK, KSPEKEE, KSPVKAE, KSPAEAK, KSPPEAK, KSPEAKT, KSPAEVK, and KVPTPEK), human NF-H tail domain is postulated to be a substrate of proline-directed kinases. The threonine-phosphorylated KVPTPEK motif suggested the existence of a novel proline-directed kinase.

L14 ANSWER 62 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:343409 CA
TI Emerging tandem **mass spectrometry** techniques for the rapid identification of proteins
AU Dongre, Ashok R.; Eng, Jimmy K.; Yates, John R., III
CS Department of Molecular Biotechnology, University of Washington, Seattle, WA, 98195, USA
SO Trends in Biotechnology (1997), 15(10), 418-425
AB A review with 68 refs. State-of-the-art techniques such as liq.-chromatog./electrospray-ionization tandem **mass spectrometry** have, in

conjunction with **database-searching computer algorithms**, revolutionized the anal. of biochem. species from complex biol. mixts. With these techniques, it is now possible to perform high-throughput protein identification at picomolar-to-subpicomolar levels from protein mixts. This article provides an overview of the techniques and methodols. available for the structural elucidation and identification of proteins and peptides from complex biol. samples.

L14 ANSWER 63 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:173334 CA
TI Direct analysis of protein mixtures by tandem **mass spectrometry**
AU Yates, John R., III; McCormack, Ashley L.; Schieltz, David; Carmack, Edwin; Link, Andrew
CS Department of Molecular Biotechnology, University of Washington, Seattle, WA, 98195-7730, USA
SO Journal of Protein Chemistry (1997), 16(5), 495-497
AB Methods to identify proteins contained in mixts. are described. The approach uses microcolumn liq. chromatog. and **automated** tandem **mass spectrometry** in conjunction with protein and nucleotide **database searching algorithms**. This approach is applied to the identification of proteins obtained by immunopptn. reactions, interaction with a GST protein fusion product, and interaction with a macromol. complex.

L14 ANSWER 64 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:146619 CA
TI Rapid 'de Novo' peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight **mass spectrometer**
AU Shevchenko, Andrej; Chernushevich, Igor; Ens, Werner; Standing, Kenneth G.; Thomson, Bruce; Wilm, Matthias; Mann, Matthias
CS Protein & Peptide Group, European Molecular Biology Lab. (EMBL), Heidelberg, D-69117, Germany
SO Rapid Communications in Mass Spectrometry (1997), 11(9), 1015-1024
AB Protein microanal. usually involves the sequencing of gel-sepd. proteins available in very small amts. While **mass spectrometry** has become the method of choice for identifying proteins in databases, in almost all labs. 'de novo' protein sequencing is still performed by Edman degrdn. Here we show that a combination of the nanoelectrospray ion source, isotopic end labeling of peptides and a quadrupole/time-of-flight instrument allows facile read-out of the sequences of tryptic peptides. Isotopic labeling was performed by enzymic digestion of proteins in 1:1 16O/18O water, eliminating the need for peptide derivatization. A quadrupole/time-of-flight **mass spectrometer** was constructed from a triple quadrupole and an electrospray time-of-flight instrument. Tandem **mass spectra** of peptides were obtained with better than 50 ppm mass accuracy and resoln. routinely in excess of 5000. Unique and error tolerant identification of yeast proteins as well as the sequencing of a novel protein illustrate the potential of the approach. The high data quality in tandem **mass spectra** and the addnl. information provided by the isotopic end labeling of peptides enabled **automated** interpretation of the **spectra** via simple **software algorithms**.

The technique demonstrated here removes one of the last obstacles to routine and high throughput protein sequencing by **mass spectrometry**.

L14 ANSWER 65 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:144502 CA
TI Application of sequential paired covariance to liquid chromatography-**mass spectrometry** data. Enhancements in both the signal-to-noise ratio and the resolution of analyte peaks in the chromatogram
AU Muddiman, David C.; Huang, Baoming M.; Anderson, Gordon A.; Rockwood, Alan; Hofstadler, Steven A.; Weir-Lipton, Mary S.; Proctor, Andrew; Wu, Qinyuan; Smith, Richard D.
CS Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA, 99352, USA
SO Journal of Chromatography, A (1997), 771(1 + 2), 1-7
AB The algorithm of sequential paired covariance (SPC) was previously reported to dramatically enhance the signal-to-noise (S/N) ratio for online sepns. combined with **mass spectrometry**. That initial study focused on a limited no. of data sets derived from the combination of capillary electrophoresis (CE) with time-of-flight **mass spectrometry** using an electrospray interface. Results from the initial study clearly demonstrated that a significant enhancement (almost two orders of magnitude) in the S/N ratio of the eluting peaks in the electropherogram could be obtained, facilitating identification of the analytes. The algorithm was applied to liq. chromatog.-**mass spectrometry** data obtained on a triple quadrupole instrument and the authors have evaluated the general applicability of the SPC approach to several types of microcolumn sepns. with **mass spectrometric** detection, including CE coupled with Fourier transform ICR **mass spectrometry**. In all the cases the authors tested, the authors found the algorithm enhanced the S/N ratios of the resulting chromatograms or electropherograms to a similar extent. This report further demonstrates the SPC approach to enhance the resoln. as well as the S/N ratio of the eluting peaks of a complex peptide mixt. While many variations of the algorithm are possible, the authors also found higher order covariance (e.g., 3rd order) is useful for eliminating coincidental noise in sequential **mass spectra**, giving the potential to ext. broad, low intensity analyte peaks. The authors also demonstrate the sequential covariance approach for enhancing the S/N ratio of **mass spectra**.

L14 ANSWER 66 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:132915 CA
TI Sequence **database** searches via de novo peptide sequencing by tandem **mass spectrometry**
AU Taylor, J. Alex; Johnson, Richard S.
CS Dep. Biochem., Univ. Washington, Seattle, WA, 98195-7350, USA
SO Rapid Communications in Mass Spectrometry (1997), 11(9), 1067-1075
AB A method is described for searching protein sequence databases using tandem **mass spectra** of tryptic peptides. The approach uses a de novo sequencing algorithm to derive a short list of possible sequence candidates which serve as query sequences in a subsequent homol.-based

database search routine. The sequencing algorithm employs a graph theory approach similar to previously described sequencing programs. In addn., amino acid compn., peptide sequence tags, and incomplete or ambiguous Edman sequence data can be used to aid in the sequence detns. Although sequencing of peptides from tandem **mass spectra** is possible, one of the frequently encountered difficulties is that several alternative sequences can be deduced from one spectrum. Most of the alternative sequences, however, are sufficiently similar for a homol.-based sequence database search to be possible. Unfortunately, the available protein sequence **database** search **algorithms** (e.g. Blast or FASTA) require a single unambiguous sequence as input. Here we describe how the publicly available FASTA computer program was modified in order to search protein databases more effectively in spite of the ambiguities intrinsic in de novo peptide sequencing algorithms.

L14 ANSWER 68 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:103558 CA
TI Spectral **interpretation** and survey analysis in ICP-MS
AU van Veen, E. H.; de Loos-Vollebregt, M. T. C.
CS Laboratory Materials Science, Delft Univ. Technology, Delft, 2628 AL, Neth.
SO Special Publication - Royal Society of Chemistry (1997), 202(Plasma Source Mass Spectrometry), 77-84
AB Software was developed for survey anal. of unknown samples using the multi-element capabilities of ICP-MS. The approach is based on the measurement and **data redn.** of full range mass scans. Anal. information is obtained about the main components, minor components, and trace elements as well as interfering compds., which are **automatically** detected and cor. Concns. are reported with an est. of the quality of the **data redn.** as the true detection limit and the RSD for the fit of the model for each element and interfering compd. In addn. to semi-quant. survey anal., the software is useful for diagnostics and for method development purposes.

L14 ANSWER 70 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 127:44137 CA
TI Automated extraction of pure **mass spectra** from gas chromatographic/**mass spectrometric** data
AU Pool, Wim G.; De Leeuw, Jan W.; Van De Graaf, Bastiaan
CS Netherlands Institute for Sea Research (NIOZ), Den Burg, 1790 AB, Neth.
SO Journal of Mass Spectrometry (1997), 32(4), 438-443
AB An algorithm is described that exts. pure **mass spectra** from gas chromatog./**mass spectrometric** (GC/MS) data. It is based on backfolding, a method described previously to enhance chromatog. resoln. in GC/MS data. The ability to ext. pure **mass spectra** was evaluated with both simulated and real GC/MS **data** and the **algorithm** was compared with two other methods described recently. The algorithm presented gives good results, even when the chromatog. resoln. is poor and the spectra are very similar. No a priori knowledge concerning the compn. of the data is required.

L14 ANSWER 75 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 126:31620 CA
TI Search of sequence **databases** with uninterpreted high-energy collision-induced dissociation spectra of peptides
AU Yates, John R., III; Eng, Jimmy K.; Clauser, Karl R.; Burlingame, Alma L.
CS Dep. Mol. Biotechnol., Univ. Washington, Seattle, WA, USA
SO Journal of the American Society for Mass Spectrometry (1996), 7(11), 1089-1098
AB The utility of the SEQUEST computer algorithm was broadened to permit correlation of uninterpreted high-energy collision-induced dissocn. spectra of peptides with all sequences in a database. SEQUEST now allows for the addnl. fragment ion types obsd. under high-energy conditions. Spectra were analyzed from peptides isolated following trypsin digestion of 13 proteins. SEQUEST ranked the correct sequence first for 90% (18/20) of the spectra in searches of the OWL database, without constraint by enzyme cleavage specificity or species of origin. All false-positives were flagged by the scoring system. SEQUEST searches databases for sequences that correspond to the precursor ion mass ±0.5 u. Preliminary ranking of the top 500 candidates is done by calcn. of fragment ion masses for each sequence, and comparison to the measured ion masses on the basis of ion series continuity, summed ion intensity, and immonium ion presence. Final ranking is done by construction of model spectra for the 500 candidates and constructing/performing of a cross-correlation anal. with the actual spectrum. Given the need to relate mounting genome sequence information with corresponding suites of proteins that comprise the cellular mol. machinery, tandem **mass spectrometry** appears destined to play the leading role in accelerating protein identification on the large scale required.

L14 ANSWER 77 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 125:264610 CA
TI A Noise and Background Reduction Method for Component Detection in Liquid Chromatography/**Mass Spectrometry**
AU Windig, Willem; Phalp, J. Martin; Payne, Alan W.
CS Eastman Kodak Company, Rochester, NY, 14652-3712, USA
SO Analytical Chemistry (1996), 68(20), 3602-3606
AB The combination of liq. chromatog. with **mass spectrometry**, particularly using electrospray as an ionization method, can result in chromatograms with a high level of background and noise. The use of background subtraction techniques or the Biller-Biemann **algorithm** to reduce this problem was of limited success. A variable selection procedure was developed that selects mass chromatograms with low noise and low background; these mass chromatograms are then combined to form a reduced total ion chromatogram (TIC) trace. This is achieved by calcg. a similarity index between each original mass chromatogram and its smoothed and mean-subtracted version. Further it is possible to reduce the no. of mass chromatograms by more than an order of magnitude without losing chem. significant information. The process results in significantly improved chromatograms and a significant **redn.** in **data**

anal. times for liq. chromatog./**mass spectrometry**. The approach is named component detection **algorithm** (CODA).

L14 ANSWER 78 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 125:213367 CA
TI Mining genomes with MS
AU Yates, John R., III; McCormack, Ashley L.; Eng, Jimmy
CS Univ. Washington, Seattle, WA, USA
SO Analytical Chemistry (1996), 68(17), 534A-540A
AB A review with ~35 refs. Searching protein and nucleotide databases with **mass spectra** data allows accurate identification of amino acid sequences.

L14 ANSWER 79 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 125:162688 CA
TI The development of a data system for a combination of liquid chromatography or capillary electrophoresis with an ion trap storage/reflectron time-of-flight mass detector
AU Qian, Mark G.; Wu, Jing-Tao; Parus, Steve; Lubman, David M.
CS Dep. Chem., Univ. Michigan, Ann Arbor, MI, 48109-1055, USA
SO Rapid Communications in Mass Spectrometry (1996), 10(10), 1209-1214
AB A data system based upon a 200 MHz transient recorder interface card in a Pentium PC **computer** is demonstrated for online anal. of microbore high-performance liq. chromatog. (HPLC), capillary HPLC and capillary electrophoresis (CE) sepns. using a fast and sensitive ion-trap storage/reflectron time-of-flight **mass spectrometric** detector (IT-reTOFMS). Under the control of a user-written program, the system is capable of conducting the data acquisition and storage for a min. of 30 min, at rates exceeding 10 Hz, of individual **mass spectra** contg. 16000 data points having 10 nsec resoln. The capability is mainly attributed to the use of a **data redn.** scheme in which only mass intensities higher than a preset threshold are saved as indexed flight-time/intensity pairs. This produces a typical redn. ratio of 30:1 in data set size, yielding faster storage with smaller file size, and permits the complete set of **mass spectra** to be held in the **computer**'s memory. In addn., the data system is capable of displaying, for real-time evaluation of the anal., each individual **mass spectrum** and the total-ion chromatogram. Further, the selected-ion chromatograms of given masses and a 3-dimensional topog. map describing a sepn. process can be rapidly generated from the collected data for the unambiguous and high fidelity identification of target analytes in a complex mixt.

L14 ANSWER 89 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 124:174880 CA
TI SPECTRA: A Spectral Information Management System Featuring a Novel Combined Search Function
AU Masui, Hideyuki; Yoshida, Mototsugu
CS Organic Synthesis Research Laboratory, Sumitomo Chemical Company, Takatsuki, 569-11, Japan
SO Journal of Chemical Information and Computer Sciences (1996), 36(2), 294-8

AB The SPECTRA collection of software as a spectral information management system for org. compd. structure detn. is described. The SPECTRA (SPECTral Research and Anal.) system suggests candidate structures for chem. compds. based on anal. of their spectra, where **mass spectra**, IR spectra, 1H-NMR spectra, and 13C-NMR spectra are possible input. The system computes the optimal matching of an input spectrum with stored **spectra** in a **database** and also retrieves the spectra of compds. that contain a substructure of the unknown compd. A novel combined search **algorithm** can be activated when two to four spectra are given as information of an unknown compd. Similarities between the input spectrum and each **spectrum** in the **database** are calcd., and the corresponding candidate compds. are ranked according to their similarity score.

L14 ANSWER 94 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 123:334241 CA
TI Error-tolerant protein **database** searching using peptide product-ion spectra
AU Bonner, Ron; Shushan, Bori
CS PE SCIEX, Concord, ON, L4K 4V8, Can.
SO Rapid Communications in Mass Spectrometry (1995), 9(11), 1077-80
AB A method for matching proteins in databases with unknown samples using data derived from peptide product ion masses is described. The power of the method is due to the speed of anal. and to the ability to locate proteins in a database even when the exptl. data show anomalies due to derivatization or post-translational modification.

L14 ANSWER 95 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 123:296706 CA
TI Comparison of softwares used for the detection of analytes present at low levels in liquid chromatographic-**mass spectrometric** experiments
AU Visentini, J.; Kwong, E. C.; Carrier, A.; Zidarov, D.; Bertrand, M. J.
CS Merck Frosst Centre for Therapeutic Research, P.O. Box 1005, Pointe Claire-Dorval, Quebec H9R 4P8, Can.
SO Journal of Chromatography, A (1995), 712(1), 31-43
AB A comparison has been made between different approaches for detecting low-level analytes in the TIC traces of sample mixts. analyzed by different liq. chromatog.-**mass spectrometric** (LC-MS) techniques. The approaches studied were contour mapping or "eagle's view" and a background treatment software, TICFilt, recently developed. Typical pharmaceutical samples including stds. and plasma contg. common drugs such as propranolol, phenothiazine, acetaminophen have been analyzed in LC-MS expts. using ion-spray, atm.-pressure chem. ionization and direct liq. introduction interfaces. The data obtained were examd. by contour mapping and treated by TICFilt to detect low level elution peaks. Contour mapping can be efficient at higher masses (>Mr 250) where the background is generally weaker but cannot always detect elution peaks at lower masses where background contribution is important. Furthermore, it cannot distinguish actual peaks from spikes which are often present in these expts. Background treatment algorithms such as TICFilt, however, can not only eliminate spikes from the TIC trace but

also offer a peak detection efficiency for unknown compds. which is const. throughout the mass range and independent of the mobile phase compn. and the ionization technique used. Furthermore, background treatment **algorithms** also provide **mass spectra** with enhanced spectral information which is important in the identification of unknown drug-related species.

L14 ANSWER 96 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 123:277006 CA
TI Mining Genomes: Correlating Tandem **Mass Spectra** of Modified and Unmodified Peptides to Sequences in Nucleotide **Databases**
AU Yates, John R., III; Eng, Jimmy K.; McCormack, Ashley L.
CS Department of Molecular Biotechnology, University of Washington, Seattle, WA, 98195-2145, USA
SO Analytical Chemistry (1995), 67(18), 3202-10
AB The correlation of uninterpreted tandem **mass spectra** of modified and unmodified peptides, produced under low-energy (10-50 eV) collision conditions, with nucleotide sequences is demonstrated. In this method nucleotide databases are translated in six reading frames, and the resulting amino acid sequences are searched "on the fly" to identify and fit linear sequences to the fragmentation patterns obsd. in the tandem **mass spectra** of peptides. A cross-correlation function is then used to provide a measurement of similarity between the mass-to-charge ratios for the fragment ions predicted by amino acid sequences translated from the nucleotide database and the fragment ions obsd. in the tandem **mass spectrum**. In general, a difference greater than 0.1 between the normalized cross-correlation functions for the first- and second-ranked search results indicates a successful match between sequence and spectrum. Measurements of the deviation from max. similarity employing the spectral reconstruction method are made. The search method employing nucleotide databases is also demonstrated on the spectra of phosphorylated peptides. Specific sites of modification are identified even though no specific information relevant to sites of modification is contained in the character-based sequence information of nucleotide databases.

L14 ANSWER 97 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 123:250387 CA
TI Polypeptide **mass spectra**
AU Fang, Huisheng; Xiang, Bingren; An, Dengkui
CS Analysis & Computer Center, China Pharmaceutical University, Nanjing, 210009, Peop. Rep. China
SO Shengwu Huaxue Yu Shengwu Wuli Jinzhan (1995), 22(4), 361-6
AB An algorithm for searching sequence-specific ions is proposed for interpretation of **mass spectra** of unknown polypeptides. This program is composed of three parts: searching, scoring and merging. The method successfully interpreted **mass spectra** of some unknown polypeptides. One of the major advantages of this program over algorithms described earlier is its scoring ability which can rank the confidence of every amino acid residue in the interpreted polypeptide. It greatly facilitates the detn. of the amino acid sequence and provides a pathway

for the application of **mass spectra** to biol.

L14 ANSWER 101 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 123:168895 CA
TI Sequence **database** searching by **mass spectrometric** data
AU Mann, Matthias
CS European Molecular Biology Laboratory, Heidelberg, D-69012, Germany
SO Microcharact. Proteins (1994), 223-45. Editor(s): Kellner, Roland; Lottspeich, Friedrich; Meyer, Helmut E. Publisher: VCH, Weinheim, Germany.
AB **Mass spectrometry** is a powerful tool in the identification of proteins. The concepts of database searching by **mass spectrometric** information are explained using PeptideSearch, a program written by the author. Searching by total mass information, searching by peptide masses and searching by a combination of peptide mass and partial sequence are presented.

L14 ANSWER 102 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 123:48602 CA
TI Error-Tolerant Identification of Peptides in Sequence **Databases** by Peptide Sequence Tags
AU Mann, M.; Wilm, M.
CS Protein Peptide Group, European Molecular Biology Laboratory, Heidelberg, D-69012, Germany
SO Analytical Chemistry (1994), 66(24), 4390-9
AB The authors demonstrate a new approach to the identification of **mass spectrometrically** fragmented peptides. A fragmentation spectrum usually contains a short, easily identifiable series of sequence ions, which yields a partial sequence. This partial sequence divides the peptide into three parts-regions 1, 2, and 3-characterized by the added mass m1 of region 1, the partial sequence of region 2, and the added mass m3 of region 3. The authors call the construct, m1 partial sequence m3, a "peptide sequence tag" and show that it is a highly specific identifier of the peptide. An algorithm developed here that uses the sequence tag to find the peptide in a sequence database is up to 1 million-fold more discriminating than the partial sequence information alone. Peptides can be identified even in the presence of an unknown post-translational modification or an amino acid substitution between an entry in the sequence database and the measured peptide. These concepts are demonstrated with model and practical examples of electro-spray **mass spectrometry/mass spectrometry** of tryptic peptides. Just two to three amino acid residues derived by fragmentation are enough to identify these peptides. In peptide mapping applications, even less information is necessary.

L14 ANSWER 103 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 122:234572 CA
TI An approach to correlate tandem **mass spectral** data of peptides with amino acid sequences in a protein **database**
AU Eng, Jimmy K.; McCormack, Ashley L.; Yates, John R., III
CS Department of Molecular Biotechnology, University of Washington,

Seattle, WA, USA

SO Journal of the American Society for Mass Spectrometry (1994), 5(11), 976-89

AB A method to correlate the uninterpreted tandem **mass spectra** of peptides produced under low energy (10-50 eV) collision conditions with amino acid sequences in the Genpept database has been developed. In this method the protein database is searched to identify linear amino acid sequences within a mass tolerance of ±1 u of the precursor ion mol. wt. A cross-correlation function is then used to provide a measurement of similarity between the mass-to-charge ratios for the fragment ions predicted from amino acid sequences obtained from the database and the fragment ions obsd. in the tandem **mass spectrum**. In general, a difference >0.1 between the normalized cross-correlation functions of the first- and second-ranked search results indicates a successful match between sequence and spectrum. Searches of species-specific protein databases with tandem **mass spectra** acquired from peptides obtained from the enzymically digested total proteins of E. coli and S. cerevisiae cells allowed matching of the spectra to amino acid sequences within proteins of these organisms. The approach described in this manuscript provides a convenient method to interpret tandem **mass spectra** with known sequences in a protein database.

L14 ANSWER 104 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 122:182450 CA

TI Method to Correlate Tandem **Mass Spectra** of Modified Peptides to Amino Acid Sequences in the Protein **Database**

AU Yates, John R., III; Eng, Jimmy K.; McCormack, Ashley L.; Schieltz, David

CS Department of Molecular Biotechnology, University of Washington, Seattle, WA, 98195, USA

SO Analytical Chemistry (1995), 67(8), 1426-36

AB A method to correlate uninterpreted tandem **mass spectra** of modified peptides, produced under low-energy (10-50 eV) collision conditions, with amino acid sequences in a protein database has been developed. The fragmentation patterns obsd. in the tandem **mass spectra** of peptides contg. covalent modifications is used to directly search and fit linear amino acid sequences in the database. Specific information relevant to sites of modification is not contained in the character-based sequence information of the databases. The search method considers each putative modification site as both modified and unmodified in one pass through the database and simultaneously considers up to three different sites of modification. The search method will identify the correct sequence if the tandem **mass spectrum** did not represent a modified peptide. This approach is demonstrated with peptides contg. modifications such as S-carboxymethylated cysteine, oxidized methionine, phosphoserine, phosphothreonine, or phosphotyrosine. In addn., a scanning approach is used in which neutral loss scans are used to initiate the acquisition of product ion **MS/MS** spectra of doubly charged phosphorylated peptides during a single chromatog. run for **data** anal. with the **database-searching algorithm**. The approach described in this paper provides a convenient method to match the nascent tandem

mass spectra of modified peptides to sequences in a protein database and thereby identify previously unknown sites of modification.

L14 ANSWER 106 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 122:172773 CA
TI A new method for the rapid deconvolution of partially resolved spectra
AU Brenton, A. Gareth; Lock, Christopher M.
CS Mass Spectrom. Res. Unit, Univ. Wales, Swansea, SA2 8PP, UK
SO Rapid Communications in Mass Spectrometry (1995), 9(2), 143-9
AB A rapid and simple method for deconvolution of partially resolved spectra has been developed. The algorithm is simple, extremely fast, and capable of deconvoluting whole spectra in real time on a fast personal computer. The technique has been evaluated using known test spectra whose resoln. has been purposely degraded. The original spectra were reproduced accurately, both in the positions and the intensities of the individual peaks they contained. The method was initially developed for translational energy spectroscopy (TES), where there is little a priori information on the spacing and widths of individual peaks in a spectrum. Examples are given for actual exptl. TES spectra of varying complexity and signal-to-noise ratios. Four key variables are required as input before the program can be executed; these have been parameterized for a range of conditions and typical values have been established. The procedure may be well suited as a rapid-pre-screening process before more elaborate techniques are used and it can be applied widely, for example to deconvolute electrospray **mass spectra.**

L14 ANSWER 118 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 120:245479 CA
TI **Computer** processing and **interpretation** of **mass spectral** information. Part IX. Generalized characteristics of **mass spectra**
AU Sukharev, Yu. N.; Nekrasov, Yu. S.; Molgachova, N. S.; Tepfer, E. E.
CS A. N. Nesmeyanov Inst. Organo-Element Compd., Moscow, 117334, Russia
SO Organic Mass Spectrometry (1993), 28(12), 1555-61
AB The formation of numerical generalized characteristics (**indexes**) of **mass spectra** by using one or two exptl. parameters (mass nos. and/or ion peak amplitudes) is suggested. The influence of the measuring error of peak amplitudes on the statistical characteristics of indexes was studied. It was ascertained by four classes of organometallic compds. [ferrocene, cymantrene, (η5-cyclopentadienyl)tricarbonylrhenium, (η6-benzene)tricarbonylchromium] that within an isolated class the value of each index is distributed by a normal law. The possibility of using **mass spectral indexes** for the identification of unknown compds. is demonstrated.

Adam 15

L14 ANSWER 120 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 120:41535 CA
TI Real-time **spectral** analysis **algorithm** for space plasma three-dimensional ion **mass spectrometers**
AU Sittler, E. C., Jr.
CS Goddard Space Flight Cent., NASA, Greenbelt, MD, 20771, USA

SO Review of Scientific Instruments (1993), 64(10), 2771-81

AB The authors have developed a fast real-time **spectral** anal. **algorithm** for space plasma three-dimensional (3D) ion **mass spectrometers** that deconvolves contributions to time-of-flight ion **mass spectra** for various ion species abundances. The algorithm is composed of a set of coupled linear equations with const. coeffs. The algorithm is implemented so that in-flight computers need only apply a predetd. no. of multiplies and adds to the spectral data. The algorithm allows run times to be short and highly predictable, can accommodate the presence of background in the ion **mass spectra**, and can be updated to adjust to calibration changes and unexpected instrument anomalies or failures. Space plasma 3D ion **mass spectrometers** have the capability of generating large vols. of data and if not compressed would produce data rates that far exceed the telemetry rate usually allocated to space plasma instruments. The real-time application of this algorithm allows one to achieve compression ratios greater than 100 for the spectral data without introducing systematic errors to the computed ion abundances. It also allows the application of other higher level data compression techniques to provide addnl. compression of the telemetry data. Finally, the algorithm can be thought of as a way to increase the mass resoln. of the ion spectrometer.

L14 ANSWER 121 OF 324 CA COPYRIGHT 2004 ACS on STN

AN 120:3992 CA

TI Peptide mass maps: A highly informative approach to protein identification

AU Yates, John R., III; Speicher, Stephen; Griffin, Patrick R.; Hunkapiller, Tim

CS Sch. Med., Univ. Washington, Seattle, WA, 98195, USA

SO Analytical Biochemistry (1993), 214(2), 397-408

AB A **computer** searching **algorithm** has been used to identify protein sequences in the Protein Information Resource (PIR) database with peptide mass information (mass map) obtained from proteolytic digests of proteins analyzed by microcapillary high-performance liq. chromatog. electrospray ionization **mass spectrometry**. A theor. anal. of the cytochrome c family demonstrates the ability to identify protein sequences in the PIR database with a high degree of accuracy using a set of six predicted tryptic peptide masses. This method was also applied to exptl. detd. peptide masses for a small GTP-binding protein, a protein from pig uterus, the human sex steroid binding protein, and a thermostable DNA polymerase. The results demonstrate that a set of obsd. masses which is less than 50% of the total no. of predicted masses can be used to identify a protein sequence in the database. For the anal. presented in this paper, a mass matching tolerance of 1 amu is used. Under these conditions, mass maps created by fast atom bombardment **mass spectrometry** and matrix-assisted laser desorption time-of-flight would also be applicable. In cases where multiple matches are obsd. or verification of the protein identification is needed, tandem **mass spectrometry** sequencing can be used to establish sequence similarity.

L14 ANSWER 133 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 118:69351 CA
TI The application of MaxEnt to high-resolution **mass spectrometry**
AU Ferrige, A. G.; Seddon, M. J.; Skilling, J.; Ordsmith, N.
CS Wellcome Res. Lab., Beckenham/Kent, BR3 3BS, UK
SO Rapid Communications in Mass Spectrometry (1992), 6(12), 765-70
AB The MaxEnt technique has previously been successfully applied to the deconvolution of electrospray **mass spectra**. The latest version of the Cambridge University software, MemSys5, has now been applied to high resoln. **mass spectra**. Initial results have shown that peaks requiring an instrument resoln. of almost 200,000 to sep. them are readily resolved by MaxEnt on data acquired at a resoln. of only 50,000. Moreover, the MaxEnt results are accompanied by quant. and realistic error bars.

L14 ANSWER 138 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 117:123775 CA
TI Pattern-based **algorithm** for peptide sequencing from tandem high energy collision-induced dissociation **mass spectra**
AU Hines, Wade M.; Falick, Arnold M.; Burlingame, Alma L.; Gibson, Bradford W.
CS Dep. Pharm. Chem., Univ. California, San Francisco, CA, 94143-0446, USA
SO Journal of the American Society for Mass Spectrometry (1992), 3(4), 326-36
AB A new strategy is reported for extg. complete and partial sequence information from collision-induced dissocn. (CID) spectra of peptides. CID spectra are obtained from high energy CID of peptide mol. ions on a four-sector tandem **mass spectrometer** with an electro-optically coupled microchannel array detector. A peak detection routine reduces the spectrum to a list of peak masses and peak heights, which is then used for sequencing. The sequencing algorithm was designed to use spectral data to generate sequence fits directly rather than to use data to test the fit of series of sequence guesses. The peptide sequencing algorithm uses a pattern based on the polymeric nature of peptides to classify spectral peaks into sets that are related in a sequence-independent manner. It then establishes sequence relationships among these sets. Peak detection from raw data takes 10-20 s, with sequence generation requiring an addnl. 10-60 s on a Sun 3/60 work station. The program is written in the C language to run on a Unix platform. The principal advantages of this method are in the speed of anal. and the potential for identifying modified or rare amino acids. The algorithm was designed to permit real-time sequencing but awaits hardware modifications to allow real-time access to CID spectra.

L14 ANSWER 142 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 117:3608 CA
TI Electrospray ionization **mass spectrometry**: deconvolution by an entropy-based **algorithm**
AU Reinhold, Bruce B.; Reinhold, Vernon N.
CS Sch. Public Health, Harvard Univ., Boston, MA, 02115, USA
SO Journal of the American Society for Mass Spectrometry (1992), 3(3),

207-15

AB A novel algorithm is discussed for extg. parent masses from spectra contg. multiply charged ions, a common feature of electrospray ionization **mass spectrometry**. The **algorithm** works with raw **data** and does not require the generation of a peak table, and is thus less sensitive to errors introduced by overlapping peaks and other problems assocd. with peak assignment. Preliminary results suggest this approach to be most effective in analyzing samples of increasing complexity.

L14 ANSWER 146 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 116:173397 CA
TI Generation of substructure identification rules using feature-combinations from tandem **mass spectra**
AU Hart, K. J.; Palmer, P. T.; Diedrich, D. L.; Enke, C. G.
CS Dep. Chem., Michigan State Univ., East Lansing, MI, 48824, USA
SO Journal of the American Society for Mass Spectrometry (1992), 3(2), 159-68
AB Software to interpret tandem **mass spectra**, entitled Method for Analyzing Patterns in Spectra (MAPS), has been developed to provide substructure information for an automated compd. identification system. This **software** consists of several program modules which manipulate **databases** of tandem **mass spectra** and substructure information, generate substructure identification rules, and apply these rules to the tandem **mass spectra** of unknown compds. to identify components of their structure. The MAPS rule generation program has been modified to generate rules based on specific combinations of spectral features that occur concertedly. False positives are drastically reduced by searching for feature-combinations that have 100% uniqueness with respect to a ref. database of compds. Recall is increased by the detn. of multiple feature-combinations indicative of the presence of a given substructure. Strategies were developed in the algorithm for the discovery of feature-combinations that avoid the computation explosion that occurs when working with a large no. of spectral features. The rules developed have the form: IF feature-combination a (FC a) or FC b, ..., or FC x, THEN substructure SSn is present.

L14 ANSWER 176 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 111:77443 CA
TI Optimization of automatically generated rules for predicting the presence and absence of substructures from MS and **MS/MS** data
AU Palmer, Peter T.; Hart, Kevin J.; Enke, Christie G.; Wade, Adrian P.
CS Dep. Chem., Michigan State Univ., East Lansing, MI, 48824, USA
SO Talanta (1989), 36(1-2), 107-16
AB A pattern-recognition/artificial-intelligence program, referred to as MAPS (Method for Analyzing Patterns in Spectra), was recently developed to identify the relations that exist between substructures and the characteristic features they produce in the spectra from **mass spectrometry** (MS) and successive **mass spectrometry** (**MS/MS**). MAPS has been extended to utilize these relationships to formulate exclusion rules as well as inclusion rules, so that the absence of recognized

substructures can be predicted as well as their presence. The potential usefulness of each MS and **MS/MS** spectral feature in such rule formulation is characterized by correlation and uniqueness factors. The correlation factor expresses the degree of correlation between a feature and a specific substructure; the uniqueness factor expresses the uniqueness of a feature with respect to that substructure. Features with high correlation factors are of most use for predicting the absence of substructures; features with high uniqueness factors are most useful for predicting their presence. Feature intensity-data have been found to improve the inclusion-rule performance and degrade the exclusion-rule performance. Criteria for optimizing the predictive abilities of both rule types are discussed.

L14 ANSWER 187 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 108:5274 CA
TI Multidimensional **computer** evaluation of **mass spectra**
AU Neudert, R.; Bremser, W.; Wagner, H.
CS BASF A.-G., Ludwigshafen, D-6700, Fed. Rep. Ger.
SO Organic Mass Spectrometry (1987), 22(6), 321-9
AB The generation of a **mass spectral** interpretation system is described that is usable both as part of a multidimensional system, and independently for the anal. of **mass spectra** only. The knowledge base is a structure-oriented **mass spectral** data collection consisting of some 42,000 spectra and topologies. The comparison of selected **mass spectral** properties such as similarity, neutral losses, and ion series of the unknown with the equiv. properties of the library spectra results in a set of corresponding structures. Subsequent substructure anal. yields a histogram of substructure frequencies contg. information about their statistical relevance. The relevant substructure set may be recombined to produce a structure proposal, as is demonstrated for 1-acetyl-2-methoxy-4-trimethylsilyloxybenzene. In a 2nd example, the relevant substructures derived by the interpretation system are used as input for the 13C-NMR substructure generator. This procedure reduces the soln. space of the structure prediction algorithm considerably. Besides the spectrum interpretation, addnl. possibilities are available. The substructure search enables, for example, a look for **mass spectrometric** reaction centers. Beyond that, substructure anal. is applicable to the detn. of structural features typical of certain combinations of neutral losses and/or characteristic fragments.

L14 ANSWER 213 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 102:192394 CA
TI Prolegomena to any future **computer** evaluation of the QCD **mass spectrum**
AU Parisi, Giorgio
CS Univ. Roma II "Tor Vergata", Rome, 00173, Italy
SO NATO ASI Series, Series B: Physics (1984), 115(Prog. Gauge Field Theory), 531-41
AB A review with 18 refs. is given on the computer applications in QCD **mass spectrum** including weak points of the various algorithms, and when possible, the way to improve them is suggested.

L14 ANSWER 254 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 92:51468 CA
TI Data-blocking cross-correlation peak detection in **computerized** gas chromatography-**mass spectrometry**
AU Bryant, William F.; Trivedi, M.; Hinchman, B., IV; Sofranko, S.; Mitacek, P., Jr.
CS Dep. Anal. Chem., Pennwalt Corp., Rochester, NY, 14603, USA
SO Analytical Chemistry (1980), 52(1), 38-43
AB A new method for the detection of mass peaks in digital data records is reported. Both a cross-correlation detection function, Dτ, (subroutine CROSS) and a data-blocking procedure based on the use of a threshold comparator and a digital clock (subroutine CTIME) can be used in processing data for a given scan. Program LOCPK combines these methods so that each is used as required by the complexity of the digital record. CROSS employs a previously unused property of Dτ to locate peaks through simple sign checking. The performance of the combined method can be verified by using an option which causes CROSS to be used exclusively. Major advantages available through LOCPK include the substantial redn. of the av. rate of **data** transmission, the **redn.** of **computer** processing time requirements, and the prodn. of **mass spectra** comparable in quality to those produced by cross-correlation anal. alone.

L14 ANSWER 279 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 86:82466 CA
TI **Computerized** data reduction of **mass spectra**
AU Hollos, Jeno
CS CHINOIN Gyogyszer- Vegyeszeti Termekek Gyara, Budapest, Hung.
SO Magyar Kemiai Folyoirat (1976), 82(10), 512-13
LA Hungarian
AB A new method is proposed for **data redn.** of **automatically** registered **mass spectra** omitting only background and small satellite peaks and resulting in clear and complete spectra. All large peaks greater than an upper threshold in each range of the spectra are stored, and the background peaks smaller than a lower threshold are omitted. Both thresholds are increased stepwise in each range directly proportional to the distance from the mol. peak up to 10 and 2.5%, resp. Some of the peaks between the 2 thresholds are also stored whose no. is in approx. inverse ratio to the no. of peaks already found and stored in the spectrum at higher mass nos. In each range the no. of stored peaks can vary from 0 to 8 according to the features of the spectra.

L14 ANSWER 285 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 85:103479 CA
TI Additional feature of the JMA-0231 GC-MS data analysis system. Private library research
AU Anon.
CS Japan
SO JEOL News (1976), 13A(1), 20-1
AB In a gas-chromatog. **mass-spectrometric** data anal. system the desired ref. data file can be assembled as a private library in a computer

memory. Ref. data can be searched for by using any of the following information items: sample name, mol. formula, integral mol. wt., exact mol. wt., whole **mass spectrum**, partial **mass spectrum**, mol. formula and whole spectrum, mol. wt. and whole spectrum, mol. formula and partial spectrum, or mol. wt. and partial spectrum. A similarity **index** of the retrieved **data** is calcd. **automatically** by using the partial spectrum.

L14 ANSWER 314 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 74:147943 CA
TI Identification of **mass spectra** by **computer-searching** a file of known spectra
AU Hertz, Harry S.; Hites, Ronald A.; Biemann, Klaus
CS Dep. Chem., Massachusetts Inst. Technol., Cambridge, MA, USA
SO Analytical Chemistry (1971), 43(6), 681-91
AB To relieve the chemist from the tedious task of manually interpreting the large no. of **mass spectra** obtained from gas chromatog. effluents, an automatic technique was developed which compares the spectrum of an unknown compd. to a large file of ref. spectra. Both the unknown and ref. spectra are abbreviated, before comparison, by selecting the 2 largest peaks in each 14 mass unit interval throughout the entire spectrum. After the **computer** preselects the most similar **mass spectra**, a similarity **index** is calcd., which represents the weighted av. ratio of the 2 spectra and is an abs. measure of the degree of match between the unknown and a particular ref. spectrum. The algorithm used is described and evaluated, and applications are presented.

L14 ANSWER 321 OF 324 CA COPYRIGHT 2004 ACS on STN
AN 68:45915 CA
TI High resolution **mass spectrometry** in molecular structure studies. XIV. Real-time data acquistion, display, and subsequent processing in high resolution **mass spectrometry**
AU Burlingame, Alma L.; Smith, Dennis Howard; Olsen, R. W.
CS Univ. of California, Berkeley, CA, USA
SO Analytical Chemistry (1968), 40(1), 13-19
AB A prototype system for acquiring high resoln. **mass spectral** data in real-time, by using a high-speed digital **computer**, is presented. The electron multiplier output of the **mass spectrometer** is digitized during a high-resoln. magnetic scan of the spectrum. The digitized raw data are suppressed by deletion of all intensities below a preset threshold and stored in the **computer** memory. The resulting data are presented on a cathode ray tube display. Then, the data can be either rejected or stored on digital magnetic tape. Subsequent **data redn.** provides extremely accurate mass measurements and reasonable intensity data. The data obtained with the system are demonstrated from the spectra of n-octadecane, 6-hydroxycrinamine, and N-acetyltetrahydropyrroles. The **data redn.** procedure, from raw **data** to final plotted output, requires 20-30 sec. of 6600 central processor time for the av. spectrum.

=> log y
STN INTERNATIONAL LOGOFF AT 19:18:11 ON 20 MAR 2004